知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

移动应用评论挖掘研究综述

张季[1] 康乐乐[1] 李博[2]
1. 南京大学信息管理学院 南京 210023
2. 中南大学商学院 长沙 410083

摘要：[目的 / 意义] 用户评论有助于开发者实现移动应用创新，通过对移动应用评论挖掘相关文献进行归纳总结，为移动应用开发和评论挖掘提供借鉴。[方法 / 过程] 利用文本分析方法，将移动应用评论挖掘相关研究归纳为评论分类、评论聚类和评论特征抽取 3 个关键主题，并基于此框架阐述该领域的发展状况。[结果 / 结论] 研究得出：评论分类方法已开始从机器学习向深度学习演变；评论聚类主要使用 K-Means 和 DBSCAN；特征抽取仍以评论的显式特征为主。未来，移动应用评论挖掘仍有 3 个问题值得探究，分别是领域依赖性、多源信息融合以及评论价值评估。

关键词：移动应用 评论挖掘 评论分类 评论聚类 特征抽取
分类号：TP391.1

引用格式：张季，康乐乐，李博. 移动应用评论挖掘研究综述 [J/OL]. 知识管理论坛，2021, 6(6): 339-350[引用日期]. http://www.kmf.ac.cn/p/266/.

1 引言

随着移动互联网的发展和移动设备的普及，移动应用（简称 APP）已经成为日常生活中不可或缺的一部分。自苹果公司 2008 年 7 月份发布 App Store、谷歌公司 2008 年 10 月份推出 Android Market（2012 年更名为 Google Play Store）之后，移动应用如雨后春笋般涌现出来。经过 10 多年的发展，Google Play Store 已有超过 345 万款应用，Apple App Store 也有近 220 万款应用 [1]，这些应用从社交媒体到新闻资讯、从商务办公到娱乐消遣、从医疗健康到学习教育、从在线购物到金融理财，涵盖了人们生活中的众多场景。2020 年，受新冠肺炎疫情的影响，人们使用移动设备的习惯向前推进了 2-3 年，移动应用下载量达到了 2 180 亿次，每个用户日均使用移动设备的时长超过了 4 小时 [2]。

移动应用的巨大需求量给 APP 开发者带来

基金项目：本文系国家自然科学基金资助项目“从驱动到扩散：基于多元方法的移动应用创新机理研究”（项目编号：72072087）研究成果之一。
作者简介：张季，硕士研究生；康乐乐，副教授，博士，博士生导师，通信作者，E-mail：lelekang@nju.edu.cn；李博，博士研究生。

收稿日期：2021-09-14 发表日期：2021-11-22 本文责任编辑：刘远颖

无限机遇的同时，也给开发者带来了巨大的挑战。第一，移动应用商店具有明显的开放性特征 [3]。在商店中，关于某一应用的功能描述、用户评论、更新文档等都是公开可见的。这意味着应用一旦发布，就面临着被模仿甚至被抄袭的风险。第二，需求分析具有典型的阶段性特征。应用程序都是针对当时的需求开发的，但在与移动应用交互的过程中，用户会不断产生新的需求。第三，市场竞争异常激烈。在特定的细分市场上，功能高度相似的应用少则数款、多则数十款，用户可以轻易地从一款 APP 转移到另一款 APP[4]。

对于移动应用而言，创新一直以来都被认为是获得竞争优势的关键来源 [5-6]。根据新颖程度，创新可分为突破式创新和渐进式创新 [7]。突破式创新是设计一个全新的产品或提出产品设计的新方法，是从 0 到 1 的过程；渐进式创新是对现有产品进行持续不断的迭代优化，是从 1 到 N 的过程。移动应用创新更多的是从 1 到 N 的过程，即对 APP 进行长期的维护和改进。不同于实体产品的创新，移动应用创新迭代非常快，如 Google Play 中的应用平均 13 天更新一次 [8]。要在如此频繁更新的情况下获得不错的市场绩效，开发者需要及时地从用户那里收集反馈。用户创新理论最先由希普尔发现并提出，该理论认为在某些行业或领域往往是用户而不是生产商提出具有创意的产品或服务 [9]。所以，这些生产商要从传统的以自己为中心的创新转向以用户为中心的创新，要为用户提供平台以激发他们的创造力 [10]。

移动应用商店的出现不仅为用户打造了一个绝佳的反馈平台，而且为开发者提供一个汲取知识的创新平台。应用商店允许用户以数字星级（从 1 星到 5 星）和开放式文本的形式发表评论 [11]，其中文本通常由标题和正文组成。在开发应用新版本时，开发者平均会使用 50% 的信息性评论 [12]。所谓信息性评论，是对提高 APP 质量或用户体验有潜在帮助的评论。然而，对开发者来说，从评论中快速筛选出信息性评论并不容易，主要原因有：①评论数量大，增长速度快。评论数量随着时间的推移会越积越多，Google Play Store 中一些热门应用每天会收到 500 多条评论 [13]，人工审阅耗时耗力。②信息性评论大约只占总评论数的三分之一 [14]。也就是说，评论中包含大量的虚假评论、不相关的评论以及非评论等垃圾评论 [15]。③评论文本是有噪声的。用户撰写的文本常常不符合语法，存在拼写错误、缩写、表情包，缺少或乱加标点符号 [16]。④不同于其他评论（如新闻评论、图书评论、影视评论），移动应用评论具有强时效性和高价值性，用户针对某一版本发表的功能错误、程序崩溃等评论，若开发者及时响应，将极大地增强用户的身份认同和使用体验。因此，诸多学者致力于探索如何自动从海量的、非结构化的、非正式的评论文本中挖掘有价值的信息，然后将其纳入软件开发环节，以促进移动应用的迭代创新。

学界围绕移动应用评论挖掘取得了众多的研究成果，已有学者对此进行了系统性综述。N. Genc-Nayebi 和 A. Abran[17] 从评论挖掘技术、领域依赖、评论有用性、垃圾评论识别和软件特征提取 5 个方面展开叙述，揭示了评论挖掘的主要研究问题。但是，该综述的分类体系较为分散，并且由于文献量不足难以对评论有用性和垃圾评论识别进行全面客观的述评。M. Tavakoli 等 [18] 针对评论挖掘技术和工具进行综述，将评论挖掘技术分为有监督的机器学习技术、自然语言处理技术和特征提取技术，并罗列了当时的评论挖掘工具。然而，其缺乏对评论挖掘技术更有深度和广度的分析和归纳。鉴于评论挖掘在移动应用创新领域具有重要的意义，且近几年 APP 评论挖掘方法已经有了新的进展，所以有必要重新梳理相关文献。

本文主要贡献如下：①筛选出利用用户评论驱动 APP 创新的相关文献；②利用文本分析方法，将相关研究归纳为评论分类、评论聚类和特征抽取三大类，以期明确该领域的发展现状；③从领域依赖性、多源信息融合以及评论

知识管理论坛，2021（6）:339-350
DOI: 10.13266/j.issn.2095-5472.2021.032

价值评估 3 个方面进行展望，为未来的研究提供参考。

2 数据来源和研究框架

2.1 数据来源

本研究英文论文选取 Web of Science 核心数据集中的 SCI-E、SSCI、CPCI 作为数据来源。在增加每个术语可能的同义词以及对检索结果分析的基础上，确定的检索式为 (TS=(“user reviews$” or “consumer review$” or “user feedback” or “user comment$”) and TS=(“mobile app$” or “mobile application$” or “app store$” or “app market$”)) or (TS = (“app review$” or “application review$”))，语言类型为 English，时间跨度为 2009-2020 年，文献类型选择 article、review 和 proceedings paper。然后，筛选出与移动应用创新相关的评论挖掘文章共 54 篇文献作为研究样本。中文论文选择中国知网全文数据库中的核心期刊作为数据来源，检索式为 (su=(‘用户评论’ + ‘用户反馈’+’用户评价’) and (‘移动应用’ +’ 应用程序’ +’应用商店’ +’ 应用市场’ +’ app’)) or (su=’ app 评论’ + ‘应用评论’)，时间跨度为 2009-2020 年。同样，筛选出与移动应用创新相关的评论挖掘文章，整理得到 13 篇文献。综合 67 篇中英文文献，对用户评论驱动 APP 创新的研究进行系统总结。

2.2 研究框架

能够表达论文核心内容的关键词或主题词的词频分布可用来研究某一领域的发展现状 [19]。笔者利用 CiteSpace V[20] 从 54 篇英文文献的标题、摘要、关键词、补充关键词中提取名词性短语，一共抽取了 226 个名词性短语。作者对统计结果作进一步处理：①删除检索词以及与检索词表达相同含义的短语（如 mobile app reviews）；②把表达相同主题的短语进行归并；③保留频次大于 3 的主题，并将主题按频次由大到小排列，如表 1 所示：

表 1 高频主题

归并后的主题	频次	原始名词性短语
信息性评论 Informative Reviews	33	valuable information(4), bug report(4), feature request(4), informative reviews(2), eliciting such critical information(2), app issues(2), user opinions(2), sudden change(2), users needs(2), effective review(1), extracting informative user reviews(1), acquiring knowledge(1), crucial information(1), potential problem(1), important information(1), bug reporting(1), different points(1), major concern(1)
方法 Approaches	28	machine learning(4), topic modeling(4), automated approach(3), various approach(2), baseline approach(2), manual analysis(2), natural language processing(2), association rule(1), automated phrase-based approach(1), ac algorithm(1), adapting information retrieval technique(1), adaptive online biterm topic modeling(1), adaptive online biterm topic model(1), abnormal topic(1), automatic topic extraction(1), adaptive topic model(1)
APP创新 App Innovation	26	actionable software maintenance request(2), evolution work(2), app update(2), release planning(2), recommended software change(2), evolution tasks(2), future maintenance(2), software evolution(2), app development(2), app software maintenance optimization(1), app maintenance(1), app development information(1), changed requirement(1), accurate evolution plan(1), application evolution(1), app software improvement(1), actionable change tasks(1)
评论挖掘 App Review Mining	25	app review mining(3), analyzing reviews(2), text analysis(2), review analysis(2), user review mining(2), mining user reviews(2), app review analysis(2), analysis of online reviews(1), data mining(1), exploiting user feedback(1), analyzing user reviews(1), addressing user reviews(1), analyzing mobile app reviews(1), analyzing informative crowd reviews(1), effective user review analytics tool(1), automatic user review mining(1), analyzing feedback(1)
开发者 App Developer	21	app developer(13), competitive environment developer(2), application developer(2), original developer(2), individual app developer(1), app developers opportunities(1)

归并后的主题	频次	原始名词性短语
评论分类 Review Classification	19	classification(4),app review classification(3), automatic classification(2), text classification(2), categorize user reviews(1), app review classification problems(1), app user review classification(1), classifying app reviews(1), accurate review classification process(1), classifying user reviews(1), associative classification(1), defining suitable classification feature(1)
情感分析 Sentiment Analysis	9	sentiment analysis(6), fine grained sentiment analysis(1), computing user sentiments(1), analyzing sentiments(1)
评估指标 Evaluation Metrics	7	high accuracy(4), average precision(2), average recall(1)
评论聚类 Review Clustering	6	clustering similar user change request(1), cohesive subgroups(1), cohesive subsets(1), cluster phrases(1), clustering algorithm(1), clustering reviews(1)
特征抽取 Feature Extraction	4	app feature extraction(1), fine-grained app feature(1), app feature(1), fine-grained feature(1)

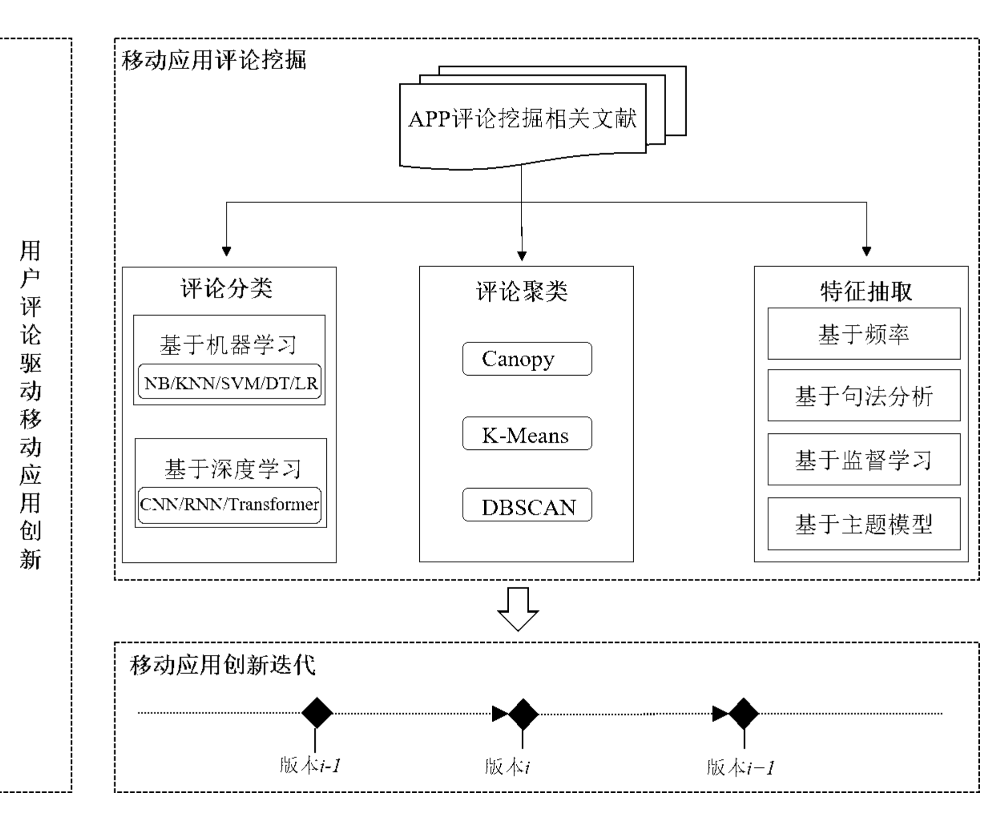


图 1 APP 评论挖掘技术路线

3 评论挖掘

3.1 评论分类

评论分类的目的不仅是要识别出有价值的评论，而且要对评论类型进行更细致的划分。通过对 Apple 应用商店中 528 条评论的人工分析，D. Pagano 和 W. Maalej 将其分为 17 个类别 [22]，其中大约一半的类别被认为与移动应用创新相关 [23-24]，如错误报告、功能请求和功能缺陷等。H. Khalid 更加关注负面评论，从 20 个 iOS 应用的 6 390 条一星或两星的评论中人工区分出 12 种类型的用户抱怨，其中功能错误、

ChinaXiv合作期刊
知识管理论坛，2021（6）:339-350
DOI: 10.13266/j.issn.2095-5472.2021.032

附加功能请求和程序崩溃等类型对开发者优化APP至关重要[25]。基于机器学习和深度学习的评论分类能够从评论中迅速识别出对开发者有用的评论类型，克服了人工分类耗时长、主观性强等缺陷。

3.1.1 基于机器学习的评论分类

移动应用评论分类的关键流程如图2所示。从图中可以看出，机器学习需要人为构建特征，有意义的特征会显著提高分类算法的性能。移动应用评论的特征可以分为语言特征和外部特征（见表2）。外部特征是指评论文本内容以外的属性，而语言特征主要包括n-grams、词性、情感。在进行评论分类时，主要利用语言特征，辅以评论元数据。常用的评论分类算法包括朴素贝叶斯（Naïve Bayes，NB）、K-近邻（K-Nearest Neighbor，KNN）、支持向量机（Support Vector Machine，SVM）、决策树（Decision Tree，DT）、逻辑回归（Logistic Regression，LR）。

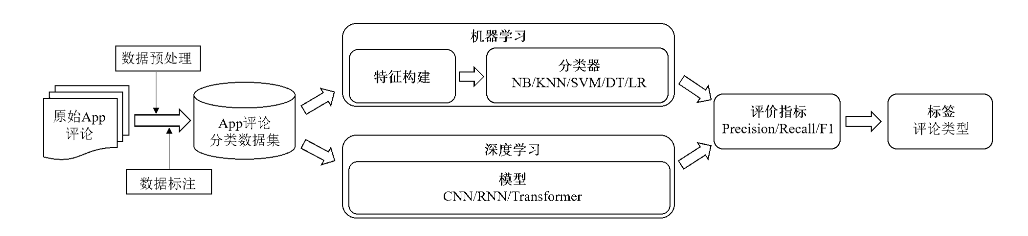


图2 移动应用评论分类的流程

表2 机器学习方法常用的特征及其描述

特征类型	特征	特征描述
外部特征	元数据	评论文本内容以外的属性，如星级评分、评论长度、评论提交时间
语言特征	n-grams	假设每个词出现的概率仅取决于该词之前的n-1个词，常用的有Unigram、Bigrams和Trigrams
	词性	不同类别的评论可能有不同的词性分布，如动词时态分布差异
	情感	计算评论的情感得分，通常正面得分为+1~+5，负面得分为-5~-1。

与单独使用文本分析、自然语言处理、情感分析和评论元数据相比，结合它们会取得更好的结果[24, 26]。W. Maalej和H. Nabil[27]进行了一系列实验来比较简单字符串匹配、词袋模型、自然语言处理（去除停用词和词形还原）、评论元数据和情感分析技术的准确率。研究发现，仅靠元数据会导致分类准确率很低，当与自然语言处理技术相结合时，分类准确率在70%-95%之间，召回率在80%-90%之间。在所有的实验中，多个二类分类器比多类分类器更准确地预测评论类型。次年，W. Maalej等[28]进一步探索，将元数据与词袋模型、自然语言处理（尤其是二元语法和词形还原）结合时，所有评论分类的准确率可达88%-92%，召回率高达90-99%。

由于有监督的方法需要人工标注训练数据，这个过程会花费大量的时间。所以在不影响准确性的情况下，主动学习和半监督学习也受到相关学者的关注。虽然主动学习和半监督学习都用到了未标注的数据，但二者的学习方式不同。主动学习是从未标注的数据中选择最易判断错误的样本交由专家标注，从而最小化训练评论分类器所需的人力，与随机选择的训练数据集相比，主动学习在多个场景下显著提高了预测的准确率[29]。然而，半监督学习是选择最不易判断错误的样本加入已标注数据。胡天媛等[30]综合分析用户评论的内容和句式结构的特点，采用半监督自学习的方式，基于有限

数量和类型的评论种子，通过循环的方式自动挖掘出体现使用反馈的 APP 软件用户评论。为了有效控制用于贬低目标应用或操纵应用排名的虚假评论，D. J. He 等 [31] 提出了一种基于 PU 学习（Positive-unlabeled learning）和行为密度（behavior density）的方法来检测虚假评论。

还有学者采用集成学习方法，以期通过聚合多个弱监督模型得到一个强监督模型。集成学习算法主要有两种：Bagging 和 Boosting。通过将朴素贝叶斯、决策树、支持向量机、逻辑回归、神经网络等不同的算法以不同的集成学习算法集成起来，大多数情况下，集成学习的性能优于单个模型 [23, 32]。

上述研究依赖于评论的文本属性，这通常会产生高维模型，并可能导致过拟合问题。因此，N. Jha 和 A. Mahmoud[33] 使用语义框架将用户评论分类为用户需求、错误报告和其他，结果表明，语义框架有助于生成更低维、更准确的模型。但是，在评论摘要任务中，基于文本生成的摘要比基于框架生成的摘要更全面 [34]。

3.1.2 基于深度学习的评论分类

深度学习相较于机器学习没有显式的特征构建过程，目前已经被广泛应用于自然语言处理问题，并在文本分类任务中取得了很好的效果。王莹等 [35] 从功能性需求与非功能性需求两个维度出发，对用户评论进行软件需求挖掘，采用 TextCNN、TextRNN 和 Transformer3 种深度学习方法，实验结果显著优于传统的机器学习方法。同样，A. Li 等 [36] 提出一种基于图卷积网络的大规模反垃圾评论模型，该模型集成了同构图和异构图来描述局部上下文和全局上下文，线上评估和线下性能都验证了该方法优于利用评论信息、用户特征和商品特征的基线模型。通常来说，深度学习在大量训练数据的情况下会有更好的表现，但在小规模的训练数据上可能并不能取得预期的效果。例如，C. Stanik 等 [37] 使用传统的机器学习方法就获得了与卷积神经网络相当的结果。当然，更复杂的模型也意味着更高的时间成本。

最后，移动应用评论分类往往牵涉训练数据类别分布不平衡的问题，这会造成分类器决策边界偏移，从而在实际应用中效果不佳。现有文献主要采用两种方式：①用代价敏感的学习方法来缓解不平衡数据的影响 [38-39]，即对不同类型的误分类设置不同的代价；②使用重采样技术来处理不平衡的类 [40-41]，即对数量多的类进行欠采样（也称为“下采样”）、数量少的类进行过采样（也称为“上采样”）。

3.2 评论聚类

评论分类是根据预定义的类别给评论分配标签，而评论聚类是将相似且没有预先划定类别的评论聚在一起。典型的聚类算法有 K-Means 和 DBSCAN，其中 K-Means 是基于形心的聚类，而 DBSCAN 是基于密度的聚类。张莉曼等 [42] 在 Word2vec 词向量模型的基础上，结合 Canopy 和 K-Means 对评论聚类，即通过 Canopy 得到聚类簇数，再运用 K-Means 得到聚类结果，该方法有效识别并聚合了用户需求。不同于广泛使用的 K-Means，DBSCAN 可以自动确定聚类簇的个数，而不需要预先指定。因此，这种方法也受到了学者的关注。L. Villarroel 等 [4] 采用 DBSCAN 算法对错误报告、新功能建议两种类型的评论进行聚类，并分别针对这两种类型的聚类簇执行优先级排序。在此基础上，S. Scalabrino 等 [43] 对评论进行了更细粒度的分类，增加了 4 类非功能性需求：安全问题报告、性能问题报告、过度能耗报告和可用性改进请求。不过，K-Means 和 DBSCAN 在移动应用评论数据集上的优劣有待进一步研究。

3.3 特征抽取

虽然评论分类或评论聚类可以从大量的评论文本中挖掘高价值的评论，但后续仍需开发者人工分析才能知道用户喜欢或讨厌的具体是哪些特征。为了解决这个问题，学者们提出了多种方法以高效地抽取 APP 特征，进而可以分析用户对这些 APP 特征的情感。笔者结合 APP 评论中特征抽取的研究现状，参照 B. Liu 对属性抽取方法的分类 [44]，将相关文献划分为 4 类：

知识管理论坛，2021（6）:339-350
DOI: 10.13266/j.issn.2095-5472.2021.032

基于频率、基于句法分析、基于监督学习和基于主题模型的特征抽取。

3.3.1 基于频率的特征抽取

基于频率的特征抽取通常先利用ICTCLAS、jieba、Standford Parser等自然语言处理工具进行词性标注，然后从标注好的语料中提取出名词、动词等，最后保留大于设定阈值的词作为候选特征 [44]。P. M. Vu 等 [45] 从原始评论中提取所有的名词和动词作为关键词，根据评论星级和出现频率对关键词进行排序，以便开发者查找与所需关键词最相关的评论。不过，单个词语只能浅显、零散地表达用户观点，而短语可以提供更完整的信息。于是，P. M. Vu 等 [46] 使用词性组合来提取用户评论中的短语，根据短语之间的相似性度量对短语进行分组，排序并监测这些分组的动态变化，从而帮助开发者获取主要的用户观点。

为了从评论中挖掘出用户高频反馈的特征，不少学者使用关联分析。这一方法的基本假设是：用户在评价APP特征时，用词是比较一致的 [47]。因此，那些频繁出现的名词或动词很可能就是APP特征。为了提高特征挖掘的效果，吕宏玉等 [48] 先利用基于句式匹配和情感倾向识别出特征请求评论，然后通过Apriori关联规则挖掘算法提取软件特征。与之不同，文涛等 [49] 利用Apriori算法提取特征后，针对每一条评论语句需要进一步识别出其中包含的＜特征词，观点词＞对。鉴于传统的频繁项集挖掘算法（如Apriori）计算量大且难以扩展，C. Gao 等 [50] 采用Eclat算法快速获得所有频率大于支持度阈值的候选短语。

3.3.2 基于句法分析的特征抽取

观点词和观点评价对象之间的评价或修饰关系往往能够通过句法关系来表征，而句法分析可以识别这些关系 [44]，从而实现特征的抽取。句法分析从语法的角度分析词语之间的关系，包括句法结构分析和依存关系分析。Z. Peng 等 [51] 使用Stanford Parser从评论的依存关系分析中提取动名词短语（动词—名词）和名词短语（名词—名词或形容词—名词），然后基于短语与主题之间的相关性，确定作为功能请求的短语。考虑到APP评论描述的内容总是与场景相关，D. Sun 等 [52] 利用评论的短语结构树和依存关系提取核心关注（kernel concern），并为每个核心关注构建聚合场景模型，帮助需求分析人员更完整、更准确地理解用户的真实意图。

3.3.3 基于监督学习的特征抽取

特征抽取任务可以转化为序列标注任务，当前主要的序列标注算法有隐马尔可夫模型（Hidden Markov Model，HMM）和条件随机场（Conditional random field，CRF）。CRF 对HMM进行了改进，打破了HMM与实际问题不符的两个基本假设——齐次马尔可夫性假设和观测独立性假设。因此，CRF在特征抽取任务中的表现更为出色，也更为常用 [53]。崔建苓等 [54] 提出基于本体和CRF融合的特征提取方法，并将深度学习Recursive Autoencoder应用于情感分析，最后形成＜特征，话题，情感词，句子，极性＞的五元组，结果表明RERM（Requirement Elicitation method based on Review Mining） 对潜在软件需求类型分类的效果良好，比ASUM（Aspect and Sentiment Unification Model） [55] 提供了更多有价值的信息。

3.3.4 基于主题模型的特征抽取

主题模型是一种生成概率模型，其目标是从文档集合中挖掘出其潜藏的主题 [56]。当前APP评论挖掘中应用最广泛的主题模型是由D. M. Blei提出的潜在狄利克雷分配（Latent Dirichlet Allocation，LDA）。LDA 主题模型利用不同文档中观测到的词来推断文档的主题分布及主题中词的分布 [57]。王欣研等 [58] 通过LDA获取用户评论主题词并运用Glove词向量相似性得到主题语义关联，然后构建出语义关联主题图谱，从而为开发者高效获取用户需求提供了新的思路和方法。近年来，学者们也提出了许多LDA的变体模型用于评论挖掘，如动态LDA[59]、自适应在线LDA[60]、E-LDA[61] 等。

除了普遍使用的LDA及其变体模型外，

ASUM[55]、非负矩阵分解 [62] 等主题模型也会被采用。另外，还有部分学者对比了不同主题模型的效果。E. Suprayogi 等 [63] 比较了 LDA 和非负矩阵分解，从主题连贯性来看，非负矩阵分解的表现更好。C. Gao 等 [64] 比较了潜在语义索引、LDA、随机投影、非负矩阵分解和基于吉布斯抽样的 LDA 模型，最终基于吉布斯抽样的 LDA 模型取得了与 AR-Miner（App Review Miner）[14] 相当的命中率，并实现了动态跟踪排名靠前的评论所反映的主要主题。

现有的主题模型大多基于 LDA 和概率潜在语义分析，但是这些主题模型对短文本的表现不佳，因为短文本会造成数据稀疏、难以识别歧义词含义等问题 [65]。为此，M. A. Hadi 和 F. H. Frad[66] 提出了自适应在线 Biterm 主题模型，有效缓解了词语共现模式稀疏的问题，可以从 APP 评论中抽取出更连贯、更高区分度的主题。

4 总结与展望

移动应用商店汇集了大量用户对 APP 的使用体验和建议，而这些反馈是开发者取得竞争优势的重要抓手，因为用户评论中包含功能缺陷、功能请求等有利于开发者优化 APP、提升用户体验的信息。笔者从评论分类、评论挖掘、特征抽取 3 个方面对相关的文献进行系统性梳理。首先，基于监督学习的评论分类仍是主流，但评论分类方法已经开始从机器学习向深度学习演变，深度学习方法在评论分类任务中的效果往往优于机器学习方法。其次，评论聚类通常作为评论分类的后续步骤，因为特定类别中的评论数量可能有数百条，通过聚类可以进一步降低开发者获取信息所付出的时间和精力。聚类算法有很多，但现有研究还没有比较不同聚类算法或算法的不同设置在移动应用评论数据集上的性能优劣。最后，有关特征抽取的文献主要集中在移动应用评论显式特征的挖掘，主题模型能够在一定程度上解决隐式特征抽取问题，但还需要专门针对 APP 评论隐式特征抽取进行研究。

未来，移动应用评论挖掘还需要深入研究的问题主要有：

（1）领域依赖性。在不同类别的应用中，词语会呈现出不同的含义，语言模式也有所不同，这使得大多数研究仅适用于特定的实验环境。例如，T. Johann 等 [67] 提出的特征提取方法 SAFE（a Simple Approach for Feature Extraction），通过人工分析应用页面和评论，确定了 18 个词性模式和 5 种句子模式，并用这些模式来提取应用页面和评论的特征。该方法对于页面维护良好的 Google Drive，精度为 87%；对于评估的 10 个应用程序，平均精度为 56%。然而，F. A. Shah 等 [68] 将 SAFE 用于 8 个不同的数据集（6 个 APP 评论数据集、1 个笔记本电脑评论数据集和 1 个餐厅评论数据集）获得的平均精度远低于论文中报告的性能。因而，APP 评论挖掘中如何实现领域迁移是一个具有挑战的研究方向。

（2）多源信息融合。一方面，不同应用商店的管理策略和用户群体存在显著的差异，使得即使是同一 APP 在不同应用商店中的用户反馈也会有所不同 [69]；另一方面，开发者不仅需要了解自身应用的优点和缺点，还要时刻关注竞争应用的长处和不足。因此，需要整合不同应用商店的用户反馈以及竞争应用的评论、产品描述和更新文档。除了从应用商店挖掘信息外，还可以收集 APP 运行时的数据。将应用商店数据和 APP 运行数据融合在一起，可以更全面地反映移动应用的状态，更准确地把握用户的需求。

（3）评论价值评估。移动应用评论的质量参差不齐，有用评论少、低价值评论多。因此，高效的评论价值评估对于 APP 开发具有积极的现实意义。当前大多数研究尚未考虑到，APP 评论价值的评估不仅仅是一个技术性问题，更是一个理论性问题。需要构建合适的价值评估体系，从多个角度对移动应用评论进行分析。具体而言，可以从评论的信息价值、时间价值、创新价值等多个维度，对移动应用评论进行恰

ChinaXiv合作期刊
知识管理论坛，2021（6）:339-350
DOI：10.13266/j.issn.2095-5472.2021.032

当的评估，以最大限度地挖掘评论的价值，更好地推动 APP 评论挖掘的演化。

参考文献：

[1] Statista Research Department. Number of apps available in leading app stores 2021[EB/OL].[2021-08-02].https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores, .

[2] App Annie. State of mobile 2021[EB/OL].[2021-08-02]. https: //www.appannie.com/cn/go/state-of-mobile-2021.

[3] BENLIAN A, HILKERT D, HESS T. How open is this Platform? The meaning and measurement of platform openness from the complementers' perspective[J]. Journal of information technology, 2015, 30(3): 209-228.

[4] VILLARROEL L, BAVOTA G, RUSSO B, et al. Release planning of mobile apps based on user reviews[C]//Proceedings of the 38th International Conference on Software Engineering. Piscataway: IEEE, 2016: 14-24.

[5] WEN W, ZHU F. Threat of platform - owner entry and complementor responses: evidence from the mobile app market[J]. Strategic management journal, 2019, 40(9): 1336-1367.

[6] MIRIC M, JEPPESEN L B. Does piracy lead to product abandonment or stimulate new product development? evidence from mobile platform - based developer firms[J]. Strategic management journal, 2020, 41(12): 2155-2184.

[7] KAPLAN S, VAKILI K. The double - edged sword of recombination in breakthrough innovation [J]. Strategic management journal, 2015, 36(10): 1435-1457.

[8] COMINO S, MANENTI F M, MARIUZZO F. Updates management in mobile applications: iTunes versus Google Play[J]. Journal of economics & management strategy, 2019, 28(3): 392-419.

[9] 黄江 . 让用户为你创造价值——用户创新理论创始人埃里克冯希普尔专访 [J]. 清华管理评论 , 2016(10): 6-11.

[10] YE H J, KANKANHALLI A. User service innovation on mobile phone platforms: investigating impacts of lead userness, toolkit support, and design autonomy[J]. MIS quarterly, 2018, 42(1): 165-188.

[11] SUSAN M M, DAVID S. What makes a helpful online review? a study of customer reviews on amazon. com [J]. MIS Quarterly, 2010, 34(1): 185-200.

[12] PALOMBA F, LINARES-VASQUEZ M, BAVOTA G, et al. Crowdsourcing user reviews to support the evolution of mobile apps[J]. Journal of systems and software, 2018, 137: 143-162.

[13] MCILROY S, SHANG W, ALI N, et al. User reviews of top mobile apps in Apple and Google app stores[J]. Communications of the ACM, 2017, 60(11): 62-67.

[14] CHEN N, LIN J, HOI S C H, et al. AR-miner: mining informative reviews for developers from mobile app marketplace[C]//Proceedings of the 36th international conference on software engineering. New York: ACM, 2014: 767-778.

[15] JINDAL N, LIU B. Opinion spam and analysis[C]//Proceedings of the 2008 international conference on web search and data mining. New York: ACM, 2008: 219-230.

[16] MCILROY S, ALI N, KHALID H, et al. Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews[J]. Empirical software engineering, 2016, 21(3): 1067-1106.

[17] GENC-NAYEBI N, ABRAN A. A systematic literature review: opinion mining studies from mobile app store user reviews[J]. Journal of systems and software, 2017, 125: 207-219.

[18] TAVAKOLI M, ZHAO L, HEYDARI A, et al. Extracting useful software development information from mobile application reviews: a survey of intelligent mining techniques and tools[J]. Expert systems with applications, 2018, 113: 186-199.

[19] 田丹 , 刘奕杉 , 王玉琳 . 热点分析类文章的文献计量分析——以词频分析方法为例 [J]. 情报科学 , 2017, 35(8): 164-169.

[20] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.

[21] THELWALL M, BUCKLEY K, PALTOGLOU G, et al. Sentiment strength detection in short informal text [J]. Journal of the American Society for Information Science and Technology, 2010, 61(12): 2544-2558.

[22] PAGANO D, MAALEJ W. User feedback in the appstore: an empirical study[C]//21st IEEE international requirements engineering conference. Piscataway: IEEE, 2013: 125-134.

[23] GUZMAN E, EL-HALIBY M, BRUEGGE B. Ensemble methods for app review classification: an approach for software evolution[C]//2015 30th IEEE/ACM international conference on automated software engineering. Piscataway: IEEE, 2015: 771-776.

[24] PANICHELLA S, SORBO A D, GUZMAN E, et al.

How can I improve my app? classifying user reviews for software maintenance and evolution[C]//2015 IEEE international conference on software maintenance and evolution. Piscataway: IEEE, 2015: 281-290.

[25] KHALID H. On identifying user complaints of iOS apps[C]//2013 35th international conference on software engineering. Piscataway: IEEE, 2013: 1474-1476.

[26] PANICHELLA S, SORBO A D, GUZMAN E, et al. Ardoc: App reviews development oriented classifier[C]//Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering. New York : ACM, 2016: 1023-1027.

[27] MAALEJ W, NABIL H. Bug report, feature request, or simply praise? on automatically classifying app reviews[C]//2015 IEEE 23rd international requirements engineering conference. Piscataway: IEEE, 2015: 116-125.

[28] MAALEJ W, KURTANOVIC Z, NABIL H, et al. On the automatic classification of app reviews [J]. Requirements engineering, 2016, 21(3): 311-331.

[29] DHINAKARAN V T, PULLE R, AJMERI N, et al. App review analysis via active learning: reducing supervision effort without compromising classification accuracy[C]//2018 IEEE 26th international requirements engineering conference. Piscataway: IEEE, 2018: 170-181.

[30] 胡甜媛 , 姜瑛 . 体现使用反馈的 APP 软件用户评论挖掘 [J]. 软件学报 , 2019, 30(10): 3168-3385.

[31] HE D J, PAN M H, HONG K, et al. Fake review detection based on PU learning and behavior density [J]. IEEE network, 2020, 34(4): 298-303.

[32] PHETRUNGNAPHA K, SENIVONGSE T. Classification of mobile application user reviews for generating tickets on issue tracking system[C]//2019 12th international conference on information & communication technology and system. Piscataway: IEEE, 2019: 229-234.

[33] JHA N, MAHMOUD A. Mining user requirements from application store reviews using frame semantics[C]//International working conference on requirements engineering: foundation for software quality. Berlin: Springer, 2017: 273-287.

[34] JHA N, MAHMOUD A. Using frame semantics for classifying and summarizing application store reviews[J]. Empirical software engineering, 2018, 23(6): 3734-3767.

[35] 王莹 , 郑丽伟 , 张禹尧 , 等 . 面向中文 APP 用户评论数据的软件需求挖掘方法 [J]. 计算机科学 , 2020, 47(12): 56-64.

[36] LI A, QIN Z, LIU R, et al. Spam review detection with graph convolutional networks[C]//Proceedings of the 28th ACM international conference on information and knowledge management. New York: ACM, 2019: 2703-2711.

[37] STANIK C, HAERING M, MAALEJ W. Classifying multilingual user feedback using traditional machine learning and deep learning[C]//2019 IEEE 27th international requirements engineering conference workshops. Piscataway: IEEE, 2019: 220-226.

[38] ZHANG L, HUANG X Y, JIANG J, et al. CSLabel: an approach for labelling mobile app reviews[J]. Journal of computer science and technology, 2017, 32(6): 1076-1089.

[39] 陈琪 , 张莉 , 蒋竞 , 等 . 一种基于支持向量机和主题模型的评论分析方法 [J]. 软件学报 , 2019, 30(5): 1547-1560.

[40] GOMAA A, EL-SHORBAGY S, EL-GAMMAL W, et al. Using resampling techniques with heterogeneous stacking ensemble for mobile app stores reviews analytics[C]//International conference on advanced intelligent systems and informatics. Berlin: Springer, 2019: 831-841.

[41] 倪瑜泽 , 彭蓉 , 孙栋 , 等 . 基于用户评论的潜在演化需求发现方法 [J]. 武汉大学学报 (理学版), 2015, 61(4): 347-355.

[42] 张莉曼 , 张向先 , 陶兴 , 等 . 面向评论语义关系的学术 APP 服务需求聚合研究 [J]. 情报理论与实践 , 2020, 43(1): 155-162.

[43] SCALABRINO S, BAVOTA G, RUSSO B, et al. Listening to the crowd for the release planning of mobile apps[J]. IEEE Transactions on Software Engineering, 2017, 45(1): 68-86.

[44] LIU B. Sentiment analysis: mining opinions, sentiments, and emotions [M]. Cambridge: Cambridge University Press, 2020: 168-171.

[45] VU P M, PHAM H V, NGUYEN T T, et al. Tool support for analyzing mobile app reviews[C]//2015 30th IEEE/ACM international conference on automated software engineering. Piscataway: IEEE, 2015: 789-794.

[46] VU P M, PHAM H V, NGUYEN T T, et al. Phrase-based extraction of user opinions in mobile app reviews[C]//Proceedings of the 31st IEEE/ACM international conference on automated software engineering. Piscataway: IEEE, 2016: 726-731.

[47] HU M, LIU B. Mining and summarizing customer

ChinaXiv合作期刊
知识管理论坛，2021（6）:339-350
DOI：10.13266/j.issn.2095-5472.2021.032

reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2004: 168-177.
[48] 吕宏玉 , 樊坤 , 杨建林 . 面向 App 用户评论的软件特征挖掘研究 [J]. 图书馆理论与实践 , 2019(7): 106-112.
[49] 文涛 , 杨达 , 李娟 . 中文软件评论挖掘系统的设计与实现 [J]. 计算机工程与设计 , 2013, 34(1): 163-167.
[50] GAO C, ZHENG W, DENG Y, et al. Emerging app issue identification from user feedback: experience on Wechat[C]//2019 IEEE/ACM 41st international conference on software engineering: software engineering in practice. Piscataway: IEEE, 2019: 279-288.
[51] PENG Z, WANG J, HE K, et al. An approach of extracting feature requests from app reviews[C]//International conference on collaborative computing: networking, applications and worksharing. Berlin: Springer, 2016: 312-323.
[52] SUN D, PENG R. A scenario model aggregation approach for mobile app requirements evolution based on user comments[M]//Requirements engineering in the big data era. Berlin: Springer, 2015: 75-91.
[53] 李志义 , 王冕 , 赵鹏武 . 基于条件随机场模型的“评价特征 - 评价词”对抽取研究 [J]. 情报学报 , 2017, 36(4): 411-421.
[54] 崔建苓 , 杨达 , 李娟 . RERM: 一种基于评论挖掘的需求获取方法 [J]. 计算机应用与软件 , 2015, 32(8): 28-33.
[55] CARRENO L V G, WINBLADH K. Analysis of user comments: an approach for software requirements evolution[C]//2013 35th international conference on software engineering. Piscataway: IEEE, 2013: 582-591.
[56] 黄佳佳 , 李鹏伟 , 彭敏 , 等 . 基于深度学习的主题模型研究 [J]. 计算机学报 , 2020, 43(5): 827-855.
[57] BLEI D M. Probabilistic topic models[J]. Communications of the ACM, 2012, 55(4): 77-84.
[58] 王欣研 , 张向先 , 张莉曼 . 学术 APP 用户在线评论主题语义关联研究 [J]. 情报科学 , 2020, 38(6): 25-31.
[59] GAO C, WANG B, HE P, et al. Paid: prioritizing app issues for developers by tracking user reviews over versions[C]//2015 IEEE 26th international symposium on software reliability engineering. Piscataway: IEEE, 2015: 35-45.
[60] GAO C, ZENG J, LYU M R, et al. Online app review analysis for identifying emerging issues[C]//Proceedings of the 40th international conference on software engineering. New York: ACM, 2018: 48-58.
[61] LIU Y, LI Y, GUO Y, et al. Stratify mobile app reviews: E-LDA model based on hot “Entity” discovery[C]//2016 12th international conference on signal-image technology & internet-based systems. Piscataway: IEEE, 2016: 581-588.
[62] LUIZ W, VIEGAS F, ALENCAR R, et al. A feature-oriented sentiment rating for mobile app reviews[C]//Proceedings of the 2018 world wide Web conference. New York: ACM, 2018: 1909-1918.
[63] SUPRAYOGI E, BUDI I, MAHENENDRA R. Information extraction for mobile application user review[C]//2018 International conference on advanced computer science and information systems. Piscataway: IEEE, 2018: 343-348.
[64] GAO C, XU H, HU J, et al. Ar-tracker: track the dynamics of mobile apps via user review mining[C]//2015 IEEE symposium on service-oriented system engineering. Piscataway: IEEE, 2015: 284-290.
[65] CHENG X, YAN X, LAN Y, et al. Btm: topic modeling over short texts[J]. IEEE transactions on knowledge and data engineering, 2014, 26(12): 2928-2941.
[66] HADI M A, FARD F H. AOBTM: adaptive online biterm topic modeling for version sensitive short-texts analysis[C]//2020 IEEE international conference on software maintenance and evolution. Piscataway: IEEE, 2020: 593-604.
[67] JOHANN T, STANIK C, MAALEJ W. Safe: a simple approach for feature extraction from app descriptions and app reviews[C]//2017 IEEE 25th international requirements engineering conference. Piscataway: IEEE, 2017: 21-30.
[68] SHAH F A, SIRTS K, PFAHL D. Is the SAFE approach too simple for app feature extraction? a replication study[C]//International working conference on requirements engineering: foundation for software quality. Berlin: Springer, 2019: 21-36.
[69] HU H, WANG S, BEZEMER C P, et al. Studying the consistency of star ratings and reviews of popular free hybrid Android and iOS apps[J]. Empirical software engineering, 2019, 24(1): 7-32.

作者贡献说明：
张　季：撰写论文初稿；
康乐乐：提出研究选题，调整论文框架，修改论文；
李　博：修改论文。

A Research Review of Mobile Application Review Mining

Zhang Ji[1] Kang Lele[1] Li Bo[2]
[1]School of Information Management, Nanjing University, Nanjing 210023
[2]Business School, Central South University, Changsha 410083

Abstract: [**Purpose/significance**] User reviews are helpful for developers to realize mobile application innovation. This paper summarizes the literature related to mobile application review mining and provides references for mobile application development and review mining. [**Method/process**] This study reviewed the researches related to mobile application review mining into three key themes of review classification, review clustering and review feature extraction by using the text analysis method, and expounded on the development status of this field according to this framework. [**Result/conclusion**] At present, the methods of review classification have begun to evolve from machine learning to deep learning; review clustering mainly uses K-Means and DBSCAN; feature extraction is still focused on the explicit features of APP reviews. In the future, there are still three issues worth exploring in mobile application review mining: domain dependence, multi-source information fusion and review value evaluation.

Keywords: mobile application review mining review classification review clustering feature extraction